

Sharing Electronic Medical Records Across Multiple Heterogeneous and Competing Institutions

Isaac S Kohane, MD, Ph.D.¹, Drs. F. J. van Wingerde¹, James C. Fackler¹, Christopher Cimino, MD³, Peter Kilbridge, MD⁴, Shawn Murphy MD, Ph.D.⁴, Henry Chueh, MD⁴, David Rind, MD², Charles Safran, MD², Octo Barnett, MD⁴, Peter Szolovits, Ph.D.⁵

¹Children's Hospital Informatics Program, Boston, MA, ²Div. Clinical Computing, Beth Israel Hospital, Boston, MA, ³Albert Einstein College of Medicine, NY., ⁴Lab. of Computer Science, Mass General Hospital, Boston, MA, ⁵Lab. for Computer Science, MIT, Cambridge, MA.

Most early reports of implemented World-Wide Web (W3) medical record systems describe single institution architectures. We describe W3-EMRS, a multi-institutional architecture, and its implementation. Thorny problems in data sharing underlined by the W3-EMRS project are reviewed.

INTRODUCTION

Despite the upheaval in the structure and economics of organized medical care, delivery of healthcare to a single patient over his or her lifetime remains a multi-institutional enterprise. Within organized healthcare delivery systems, such as managed care organizations, several distinct institutions may participate in the care of the patient. In emergency care, or for specialized referral services, the patient may obtain care outside their usual healthcare network. In our increasingly mobile society, change in employer or employment location and consequent change in healthcare plan and/or providers is hardly an uncommon event in a patient's lifetime. Safe, comprehensive and cost-effective patient care depends on the ability to obtain an accurate history of the patient's health care. In the absence of this information, tests may be repeated or prior results ignored, idiosyncratic reactions may be misreported, and drug dosage details may be miscommunicated. In the emergency care of an unconscious patient, lifesaving information may be unavailable. The clinical information relevant to a particular patient from multiple providers and institutions that may have had care responsibility over the patient's lifetime should therefore be accessible in a timely manner. Whereas Electronic Medical Record Systems (EMRS) have begun to be implemented on a large scale to provide efficient access to clinical data at individual sites, the technologies and policies of inter-institutional EMRS data sharing are currently only in the design or prototype phase.

We have implemented an architecture called W3-EMRS that enables sharing of patient data across multiple institutions and automatically collates the patient data arising from these multiple sites. This architecture takes advantage of emerging standards for communicating clinical data, and the rapidly evolving capabilities of W3. This paper motivates and describes our architectural decisions and addresses

several of the many thorny problems that even a technically successful, data sharing architecture must grapple with. Our concerns about the need for resolution of these issues has caused W3-EMRS to remain in a purely experimental and demonstration state. That is, all the databases it accesses represent small numbers of patients for which all patient and physician identifiers have been replaced and all narrative text has been manually reviewed and edited to remove any identifying information. Further, each database has been loaded into its own database, separate from any hospital system.

In our work to date, we have encountered four particular challenges: display of data obtained from different institutions, selection of patients in the absence of a national identification system, vocabulary translation, and authorization of access to records outside an institution. For developers of inter-institutional data sharing efforts, awareness and sensitivity to these issues is likely to be central to the success of their efforts. The W3-EMRS architecture permits experimentation with different solutions to these challenges but does not necessarily provide them. In effect, W3-EMRS overcomes many of the technical problems of sharing clinical data only to reveal more clearly differences in clinical practice, documentation, protection of privacy and use of medical terminology.

The World Wide Web (W3) is the collection of software protocols that allows users of computers throughout the global network of interlinked computers—the Internet, to access W3 files, typically formatted in the HyperText Markup Language (HTML), on any other machine on the Internet so long as that machine can respond to the W3 communication protocol, the HyperText Transfer Protocol (HTTP).

We implemented, in November 1994, our first working W3 system for accessing EMRS. For this first early demonstration system, we used the database of a pediatric EMRS—The Clinician's Workstation (1)—which includes access to the Children's Hospital Integrated Hospital Information System (IHIS) data repository. The data repository uses an Oracle database management system. One version of this system is in operation running against the entire Children's IHIS, but within the hospital's network

"firewall". Initial clinician response to this early W3 results reporting system has been positive (2). Furthermore, congruent with the experience of early implementors of W3-based EMRS (W3BE) (3, 4) clinicians have been able to use the prototype with no or minimal training.

The first W3-based results reporting system at Children's, and most of the aforementioned W3BE depend on the particulars of their local, existing EMRS. By "existing EMRS", we mean the system that is primarily responsible for managing the information requirements of healthcare providers. Therefore, to the extent that a W3BE's functionality depends on the particulars of an existing EMRS, it does not address the problems of sharing data between heterogeneous systems. The heterogeneity of these EMRS occurs at many levels including their database manipulation language, their information model, their local vocabulary and the policies and organizational structures of each institution. The W3-EMRS architecture allows the use of W3 browsers, and other user interface technologies to share data across multiple institutions despite their heterogeneous particularities. In the following sections we motivate and describe the architecture, and then outline its implementation. Subsequently, we review specific challenges and unresolved issues in this and similar efforts in multi-institutional clinical data sharing efforts.

ARCHITECTURE

A common computer science technique for developing and maintaining complex data exchange and transformation tasks is to develop a series of abstraction layers. Ideally, each abstract layer provides a distinct service which can be implemented independently of all the other layers. We applied this technique to the problem of moving data from multiple data repositories to a consistent user interface. The W3-EMRS abstraction layers include: 1) A common information model, called the Common Medical Record (CMR). 2) A shared set of conventions for visual presentations of the clinical data represented in the CMR. 3) The Screen Rendering layer which is the set of programs that implement the abstract functions of the visual presentation layer as user interfaces on the clinician's computer.

The CMR is designed to bridge the divergent information model of each institution. For example, in some EMRS, the problem list is directly linked to the clinical note that describes each problem whereas in others the problems and clinical notes are stored separately and can only be indirectly linked by date. Sharing data between two EMRS typically involves converting data from the information model in one institution to that in another. The number of such translations grows approximately, as the square of the number of EMRS involved. However, if there were a single standard information model, then it could serve as an intermediary information model for all EMRS

and therefore the number of distinct translations would only grow linearly with the number of EMRS.

The development of a shared information model—the CMR—has been a central task as it is a prerequisite for data sharing. The definition process was driven by our medical task domain, obtaining patient data during an emergency room visit at any of our hospitals. Our initial task was to rank in priority those medical data objects which would be the most useful to share for the acute management of the patient in the emergency room. These included active problems, clinical notes, medications, allergies, patient demographics, and the visit history. For each data object, we had to agree upon its attributes (e.g. some of the problems CMR object include problem name, vocabulary used, date first noted, comments, document describing the problem) based both on what was required for the task domain and by what was available in the existing EMRS of each of the hospitals. As could be expected, arriving at a consensus set of attributes was challenging, particularly as a close correspondence between these attributes and those of the existing EMRS would both simplify the implementation of a translator from the existing EMRS and also might make the translation process more efficient and therefore, swifter. For example, if the existing EMRS did not store the problem corresponding to each clinical note, the translation program would have to compute this correspondence based upon the dates of the problem and clinical note. In retrospect, the specificity of the emergency room task domain was particularly helpful in moving us to a consensus on the CMR.

Several options were available to implement the CMR, ranging from creating a relational database with data tables that correspond to each data object in the CMR to the definition of an arbitrary shared message format. We selected to implement the CMR within a messaging format but chose to exploit the larger consensus represented by the Health Level Seven (HL7) messaging standard. HL7 is one of the more popular interchange formats for communication between clinical information systems. HL7 communicates data as a sequence of defined ASCII characters which is hierarchically organized into segments, fields, and components. Although future versions may describe a full clinical information model, the HL7 information model remains underspecified and incomplete. That is, it allows multiple alternate messages for communicating the same data and also it does not have a set of standard segment or component formats for all the data types one might find in an EMRS. Consequently, where HL7 has defined a particular attribute of a CMR object, we adopted it. Where there are several HL7 structures for representing the same datum, we agreed which one was to be used. Where an HL7 structure for a CMR attribute was absent, we used the HL7 mechanisms for defining additional message segments to represent that attribute. Also, we had to agree on the format of standard CMR transactions, such as the HL7 format to issue a query for the problem list over a specified date range.

The Visual Presentation layer is the second abstraction layer designed to provide independence from the particular presentation conventions of an existing EMRS. It defines a set of archetypal clinical data presentation motifs such as flowsheets, timeline graphs, annotated pictures and how the elements of these presentations correspond to the CMR. It describes whether data from multiple institutions will be shown in one time ordered list, in a table with a separate column for each institution or in multiple separate pages. The Visual Presentation layer also defines permitted user interactions with these Visual presentation elements. In a table of clinical notes, and problems, it may be specified that if the clinician selects a particular clinical note, the entire text of the clinical note is retrieved and presented to the clinician.

The topmost layer in the W3-EMRS architecture is the Screen Rendering layer. This can be implemented in any technology that uses abstractions of the Visual Presentation layer to guide the creation of the corresponding displays and functions on the clinician's computer's screen. Although several technologies could be used for this purpose (e.g. Visual Basic or Powerbuilder) we have focused on W3 presentation technologies for the reasons stated above.

To implement the W3-EMRS architecture, a set of translator programs have to be implemented to mediate the set of transformations between the existing EMRS, the abstraction layers. The only translator that needs to be rewritten for each distinct EMRS is the one that translates queries against the CMR into queries in the database manipulation language of the existing EMRS and then returns the results of the query in the agreed conventions of the CMR. This is referred to as the Site Server program.. We have implemented a W3-EMRS architecture that provides these translators and have written Site Server translators for scrubbed databases extracted from three existing EMR (Children's CWS, (1) the Beth Israel's Clinical Information System (5), the MGH EMR (6)) with quite divergent information models. This implementation is described below.

IMPLEMENTATION

In the description below, it should be clear that the databases referred to are the aforementioned scrubbed databases extracted from each hospital's existing EMRS and not the actual production systems at each site.

The implementation is most simply outlined by describing how a query is executed. Selecting a button or data element displayed on the browser generates a query that is transmitted by the HTTP server to a program called the "Agglutinator." The Agglutinator broadcasts, using the HTTP protocol, this query as an HL7 message to all the sites that might have some relevant patient information (i.e. the three sites of our EMR "Collaborative"). At each existing EMRS site, upon receipt of the HL7 query, the Site Server translates the query into the data manipulation language of the existing EMRS. After the results of

the query are obtained, they are repackaged as an HL7 message that contains all the specified elements of the CMR and sent via the HTTP protocol back to the Agglutinator. The Agglutinator program then unpacks all the HL7 messages from each site server and builds data-type specific Visual Presentation objects (e.g. a table for the problem list with a different column for each hospital site) (7).

Once, the Visual Presentation objects have been generated, they are translated into the Screen Rendering layer which in this instance is HTML. These pages are sent by the Agglutinator to the HTTP server which then forwards them to the web browser where the query originated. If the Visual Presentation object for the problem table specifies a query that is associated with each data element in the table (e.g. what are the clinical notes associated with the problem) then this query is embedded as hidden fields in the HTML pages so that the query is sent to the Agglutinator if the clinician selects a problem. In this way, the clinician can review the interlinked records of a patient across multiple hospitals.

In the implementation, described above, the display of clinical data is uniform and independent of the underlying existing EMRS. It is also fixed by the programmed design of the Visual Presentation objects in the Agglutinator. We therefore have implemented a program called WYSIWYG HTML Authoring for Medicine (WHAM) (8) that allows non-programmers to generate clinical presentation specifications using a "drag and drop" interface similar to those of popular software for drawing.

In the implementation described above, the physical location of the components of the W3-EMRS architecture will determine performance of the system and several security considerations. Clearly, the W3 browser is located wherever the clinician happens to be, in this case the Emergency Room where the patient is being seen. Also, because most institutions want to have tight control over any process that accesses their EMRS, the Site Server is located at the site of the EMRS it accesses. There is much more flexibility in the location of the Agglutinator. Regardless of its location, the Agglutinator can access any specified EMRS from any point on the World Wide Web. It is unclear how many Agglutinators would be required to serve large numbers of institutions, or whether the Agglutinator is best located at every site or at a few central sites.

PROBLEMS IN DATA-SHARING

Shared Data Presentation

In the presentation of data from multiple institutions, it is tempting to integrate the data from all the EMRS into displays that collate similar data into single tables or graphs. It has become clear, even with the restricted data sets that we have used in our experimental multi-database demonstration that there are substantial differences in semantics, clinical practice and documentation style that make simple collation misleading and potentially threatening to

patient care. For example, if the Agglutinator receives an active medication list from each of the three institutions of the Collaborative, it is possible to generate a single table showing all the medications, sorted by time. However, this table might be misleading. Whereas in some EMRS, the start and end of each medication prescription might be encoded, in most others only the start or prescription date might be noted. This might lead to a dangerous misunderstanding. Similar difficulties arise with the other data types of the CMR. If in one institution, problems are added to the problem list only once, when they are first noted (e.g. asthma) and in others for each episode associated with that problem (e.g. asthma is added to the list for each exacerbation), this may mislead clinicians as to the chronicity or relevance of the problems to the patient condition. Also, we have found that in one institution, problem lists tend to include findings whereas in another they are principally used as lists of diagnosed or suspected pathological conditions. To address this problem, we are investigating a range of solutions to minimize the potential for misleading presentations including: separate display tables for each responding EMRS, or integrated tables labeled in various ways to distinguish the institutional source of each datum (e.g. different columns for each institution, colored icons corresponding to each institutions). In the future, some heuristic collation techniques may be used. There are other data types, such as clinical documents, which tend to carry along enough of the context in which they were created (e.g. in the body of the narrative text of a clinical note) so that the likelihood of mistaken interpretation is diminished.

Standardized Vocabulary

In the E.R. task domain, the use of a single standardized vocabulary is perhaps not quite as pressing as in other domains. If the problem list at MGH includes Grave's Disease and the problem list at the Beth Israel Hospital includes Thyrotoxicosis, most clinicians will understand how the problems are related. Nonetheless, if in one EMRS there is a problem of "Palpitations" five years ago and in another "Arrhythmia" two months ago, it would be useful to know whether the palpitations signified arrhythmia in the first EMRS. Also, linkage of W3BE to W3 decision support tools (e.g. the On-line Mendelian Inheritance of Man and MEDLINE) only works well if the vocabulary of the EMRS can be mapped reliably to the vocabulary of the W3 resources (e.g. MESH for MEDLINE). Furthermore, if the W3-EMRS architecture is to be used for aggregate queries across institutions, using the CMR as a common interchange format, then it becomes essential to use a standardized vocabulary so that, for instance, all patient's with Grave's Disease can be correctly identified. Several national and international efforts are underway to enable mapping between multiple vocabularies, most notably in this country, the Unified Medical Language System (UMLS) Metathesaurus (9), sponsored by the National Library of Medicine.

We are investigating implementing UMLS-based translation procedures as part of the Site-Server functions within the W3-EMRS architecture. Some of the clinical vocabularies for the data objects in the CMR are already the same in all the EMRS of the Collaborative. Specifically, the names of medications and their NDC codes are all drawn from the database of the same commercial vendor. For other attributes such as the problem name, each member of the Collaborative has their own vocabulary. To present problems in a single, standardized terminology, we have to provide the means to map terms in each local vocabulary to a standardized vocabulary. If the local vocabulary is in the UMLS, then the UMLS Metathesaurus can be used to guide automatic translation to a target vocabulary that is also within the UMLS. Of our three EMRS, only one, the MGH Costar vocabulary is currently being added to the UMLS Metathesaurus. If the local vocabulary is not in the UMLS, then each Site Server must be able to provide a mapping function from the local vocabulary to one of the UMLS standard vocabularies.

As the benefits for shared, standardized vocabularies becomes clearer from multi-institutional projects, such as this one, we anticipate that many more EMRS will adopt standardized vocabularies as their local vocabulary to avoid the cost of customized translation.

Patient Selection

Reliable selection of a patient's record is a known problem for even single institution EMRS. Name and date of birth is insufficient to reliably and uniquely identify patients. This problem is one of the reasons why most EMRS create a unique identifier for each patient seen. Even so, this does not prevent errors such as multiple ID's per patient. With multi-institutional record access, the problem is larger. In the absence of a national health identifier, we are reduced to using collections of patient attributes to identify the patient. The attributes supported by all the EMRS of the Collaborative include name, date of birth, gender, address, insurance plan and telephone number. Each Site Server can respond with an HL7 message describing all the patients matching the attribute values sent in a query. The E.R. clinician can then specify which of the patients is the desired one or reformulate the initial query. Potential pitfalls include overspecifying a query and thereby excluding the desired patient (e.g. using an out of date address) or selecting an incorrect patient from a list retrieved by an underspecified query. These problems are well known to implementors of single-institution EMRS and their solution is not tractable by computational techniques alone. Instead, the solution typically involves institutional, organizational and workflow changes. Similarly, the solution of these problems at the inter-institutional level is likely to involve the creation of regional information infrastructure that includes a regional, cryptographically protected (10), master patient index or directory.

Confidentiality Across Multiple EMRS

An obvious potential problem of W3-EMRS is the opportunity to abuse access privileges to invade a patient's privacy. Although similar concerns exist within any single EMRS, they are more acute here because W3-EMRS uses a publicly accessible network—the Internet—as its communications medium. Fortunately, because there is such a large public interest in keeping Internet communications secure (e.g. to protect credit card transactions) the cryptographic protection of communications over the Internet is continually being challenged and upgraded. Nevertheless, it remains the case that no matter how good the cryptographic protection provided, such systems remain vulnerable to "insider" abuse. Therefore institutions contemplating sharing data must first have a shared policy for data security, authentication and disciplinary action in case of breach of this policy. It may be that in the coming year, the U.S. Congress will enact medical privacy legislation that addresses some of the issues relevant to data sharing. In the interim, the Collaborative has drafted a Common Confidentiality Statement to serve as policy for our own efforts. This Statement includes several requirements of any institution participating in W3-EMRS. Among these are: 1) Each E.R. will have one provider at all times designated as being responsible for accessing information from W3-EMRS. 2) The designated E.R. provider will use both a hardware token and a hand-entered password to authenticate themselves to W3-EMRS. 3) Access to W3-EMRS will be limited as a matter of policy to the time they are in the E.R. 4) Digital signatures, cryptographically authenticating both the E.R. site and each site server will be required. The digital signatures will have to be issued by a mutually agreed upon trusted certifying authority. 5) Patient consent for record access.

SUMMARY

The first generation of W3BE were single institution implementations that have already shown significant user acceptance. The W3-EMRS project of the Boston Collaborative represents an initial set of experiments in implementing a second generation W3BE, one that takes full advantage of the geographic breadth of the Web to allow access to the longitudinal patient record as it is distributed across multiple sites, providers and institutions. Other groups, notably those of the CERC/ARTEMIS project (11) are developing second generation W3BE. Our experience with W3-EMRS suggests that if the institutions, accessed by a second generation W3BE, have truly heterogeneous EMRS, then all of them will have to resolve the challenges we have noted above. Each of these challenges are substantial. Fortunately, there is a large body of medical informatics research in these areas and we are optimistic that we will be able to leverage these prior efforts to meet their challenge. We have found that the implementation of W3-EMRS and its ability to

dynamically collate a uniform view across heterogeneous EMRS provides a useful laboratory for testing our solutions to these problems. Finally, we have found the W3-EMRS project to be an important focus for the collaboration of informatics researchers interesting in advancing the state of the art of EMRS. To this end, none of the W3-EMRS code or design is proprietary. Members of the informatics community interested in collaborating on this effort may contact the authors or visit our web site at <http://www.emrs.org/>.*

REFERENCES

1. Kohane IS. Getting the Data In: Three-Year Experience with a Pediatric Electronic Medical Record System. In: Ozbolt JG, ed. Proceedings, SCAMC. Washington, DC: Hanley & Belfus, Inc, 1994:457-461.
2. Kohane IS, Greenspun P, Fackler J, Cimino C, Szolovits P. Building National Electronic Medical Record Systems via the World Wide Web. *Journal of the American Medical Informatics Association* 1996;3(3):191-207.
3. Willard KE, Hallgren JH, Sielaff B, Connelly DP. The deployment of a World Wide Web (W3) based medical information system. In: Gardner RM, ed. SCAMC. New Orleans, Louisiana: Hanley & Belfus, Inc, 1995:771-775.
4. Cimino JJ, Socratous SA, Grewal R. The informatics superhighway: Prototyping on the World Wide Web. In: Gardner RM, ed. SCAMC. New Orleans, Louisiana: Hanley & Belfus, Inc, 1995:111-115.
5. Bleich H, Beckley R, Horwitz G, et al. Clinical computing in a teaching hospital. *New England Journal of Medicine* 1985;312:756-764.
6. Barnett GO, Castleman P. A time-sharing computer system for patient-care activities. *Comput. and Biomedical Research* 1967;1:41.
7. Wingerde FJv, Szolovits P, Kohane IS, et al. Using HL7 and the World Wide Web for unifying patient data from remote databases. In: Cimino J, ed. SCAMC. Washington, DC: Hanley & Belfus, Inc., 1996.
8. Hinds A, Greenspun P, Kohane I. WHAM! A forms constructor for medical record access via the world wide web. In: Gardner RM, ed. SCAMC. New Orleans, LA: Hanley & Belfus, Inc, 1995:116-120.
9. Lindberg D, Humphreys B. The Unified Medical Language System (UMLS) and computer-based patient records. In: Ball M, Collen M, eds. *Aspects of the computer-based patient record*. New York: Springer-Verlag, 1992:165-175.
10. Szolovits P, Kohane I. Against simple universal health identifiers. *Journal of the American Medical Informatics Association* 1994;1(4):316-319.
11. Jagannathan V, Reddy YV, Srinivas K, et al. An overview of the CERC ARTEMIS Project. In: Gardner RM, ed. SCAMC. New Orleans, Louisiana: Hanley & Belfus, Inc., 1995:12-16.

* This research was supported by the National Library of Medicine (U01 LM05877-01, LM05854, LM4-3512), the Agency for Health Care Policy and Research (HS 08749), and in part by Sun Microsystems and Oracle Corporation.